Audio Engineering Society

# Conference Paper 18

Presented at the International Acoustics & Sound Reinforcement Conference

2024 January 23-26, Le Mans, France



# Requirements and Solutions for Audio Networking in Sound Reinforcement Systems

Henning Kaltheuner[1], Genio Kronauer[2], Morten Lave[3] and Etienne Corteel[2]

[1]Consultant for d&b audioechnik GmbH & Co. KG
[2]L-Acoustics, Marcoussis, France
[3]All Access Consulting for Adamson Systems Engineering

Correspondence should be addressed to Henning Kaltheuner (henning@kaltheuner.com)

## ABSTRACT

Market demands for new business models and for multifunctionality of venues with fully digitalized technical structures foster a fast development of fully networked AV and sound reinforcement systems. At the same time the ProAV industry is experiencing a phase of disillusionment about the suitability and future proofness of network solutions for professional applications. While most established media network solutions build on standard legacy Ethernet it turns out that this technology does not meet requirements of large-scale sound reinforcement systems regarding audio performance, usability, reliability, and scalability. Leading sound reinforcement manufacturers have tackled this fundamental problem and since 2016 developed a collaborative approach named MILAN which is based on open deterministic IEEE AVB resp. TSN network technology. The paper describes the user and market requirements that lead to this decision and explains how MILAN and its underlying technology fulfils even the highest demands for audio performance, reliability, ease of use and scalability.

## 1 Introduction

Pro Audio Markets are exploring a very fast trend towards digitalized venues and systems which enable performing a broader variety of programs as well as a wide range of additional digital services for maintenance and support. While in early days of Audio Networking mainly the advantages for system cabling regarding cost, effort and flexibility led to decisions for networked products and systems [1][2], the trend for full system digitalization creates a necessity for manufacturers and integrators to make networked audio systems reliable, useful, efficient, and integrable with other infrastructure of buildings and event systems.

This paper investigates the specific technical requirements for networked Sound Reinforcement systems as well as the most critical user needs related to networking in Sound Reinforcement contexts. It explains which drawbacks media network solutions gain from the non-deterministic character of legacy Ethernet and how latest IEEE AVB and TSN Ethernet technologies form a basis for new standards with outstanding audio performance and usability.

## 2 General System Architectures

Sound Reinforcement Systems mostly consist of several loudspeakers which are arranged in certain functional groups such as [Left], [Right], [Subwoofers], [Out-Fills] (see Figure 1). Within such functional groups the loudspeaker often consists of mechanically separate elements, which are arranged in a certain way to provide certain characteristics such as horizontal and vertical dispersion, a certain waveform to be radiated or even electronically steerable dispersion by manipulating the time and phase relation between the elements with dedicated filters in DSP processors.

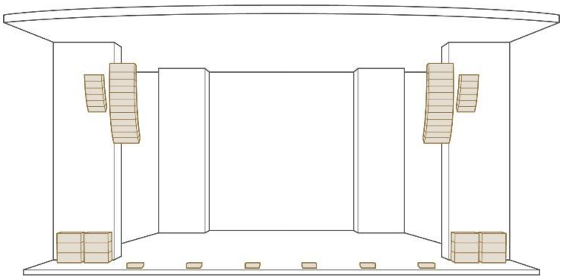

Figure 1: Functional structure of a Sound Reinforcement system

The amplification for the loudspeaker is either provided within the individual cabinet as a 'Powered Loudspeaker' or by an external amplifier unit. Because we want to discuss Audio Networking in general terms for Sound Reinforcement regardless of a specific system architecture, we take the most critical case as a criterion where each loudspeaker element is a separate network node.

## 3 Technical requirements

This section lists the core technical requirements that must be fulfilled by a network technology for Sound Reinforcement Systems.

**System Scale**
Modern sound reinforcement systems commonly comprise several hundreds of loudspeakers and related amplifiers as well as DSPs. One such example is the Metallica M72 tour that comprised 288 Meyer Sound Panther individually processed and amplified Line Array elements on a single network. **1000 network nodes** can be seen as a realistic requirement with some headroom.
The same components are also deployed into much smaller systems comprised 10 to 20 network nodes. Therefore, a network technology must scale between these extremes without causing cost and workflow overheads at either end of the range.

**Latency**
Total latency from a microphone to a loudspeaker should be in the order of 5 to 15ms depending on the size of the installation. As a result, a single network source node to sink node should be in the range of **1ms** to accommodate for at least two such paths and ADC and DAC delays.
Such network transport latency must have a guaranteed upper bound with zero probability of being higher to avoid audio dropouts in the absence of physical layer errors.

**Synchronicity**
Network nodes should be time aligned to guarantee that any acoustic or electric summation does not create any sound quality degradation. Audio time offsets among nodes are mainly an effect of errors in the synchronization of the wall clock and hence the quality of the method used for distributing the wall clock. Audio band and high frequency jitter can be avoided using modern Phase Lock Loop (PLL) designs and other methods [4]. However, there may be remaining constant time offsets or low frequency jitter, often referred to as wander [3].
The requirement may be expressed as a **maximum of Phase Deviation** that can be tolerated between individual loudspeaker elements. A maximum Phase Deviation of 10° @ 20kHz between two signals in the network would result in a loss of level of no more than 0.1 dB at 20 kHz in case of pure electric summation. This is well below the JND for spectral distortions in frequency responses of systems [5]. Expressed as a time offset, this becomes a maximum synchronicity error of roughly **1µs.**

**Dependability**
Sound Reinforcement systems are highly critical with regards to reliability. First, this has security aspects because modern speaker systems are often very powerful and could in case of errors exhibit harmful SPL to audiences and artists.
More generally the value of a loudspeaker system for performances is determined by its dependability for delivering high quality of sound even though these systems are mostly very complex and often designed and set up within a short time.
This means that for Sound Reinforcement systems usually no compromises in reliability and dependability are acceptable. System and Sound Engineers must be able to rely fully on the integrity of a system and its operational status. Whenever Sound Reinforcement systems are based on a networked audio and control infrastructure, the network becomes a very critical factor to the overall dependability.

**Controllability**
Controllability generally means that devices in a Sound Reinforcement system are network nodes both for audio as well as for control. It is of critical relevance for the network architecture and the device design that control data and audio data can coexist without interferences in one physical network. When

this is not the case, separate network ports must be provided for audio and control and switches must provide separate networks for the different types of traffic.

Device control is subject to different levels of criticality. Some control data can be treated as any best effort data traffic, some is highly critical regarding latency and dependability. These levels can be described as follows:

- System configuration: recalling presets, network setup, remote maintenance, firmware update,
- Static parameters: EQ, DSP parameters, tuning,
- Dynamic parameters: localization, dynamic beam steering, telemetric functions.

| Control type | Criticality | Latency |
|---|---|---|
| System Configuration | Low | Best effort |
| Static parameters | Low-Mid | 10-50ms for smooth operation |
| Dynamic parameters | Very High | 1ms |

Table 1: Control data types and associated constraints

## 4 Requirements for usability

**Workflow efficiency**
Sound Reinforcement Systems are often very complex systems with hundreds of components. Often they need to be designed, assembled, set up, tuned and managed within a given time in the operational processes of a venue or a production. The responsible System Engineers are constantly aiming for the highest efficiency in the related workflows and especially for those tasks, which only enable the operational status of a system. A signal distribution structure between amplifiers, powered loudspeakers, matrixes and sources should ideally cause the lowest possible effort to set it up and ensure its reliability. This counts also for installations where more and more often systems must be modified daily according to the changing requirements of productions.

**Simplicity**
Efficiency in Workflows also calls for simplicity regarding the technologies that are used within systems. In best case network structures are easy to handle, do not require deep understanding of low layers of technologies and devices and the number of rules to follow is low. Leading manufacturers of Loudspeaker systems, such as d&b Audiotechnik, L-Acoustics, and Adamson System Engineering, claim that they aim for their users to require far less than 10% of their entire time for network system design and deployment of the networked signal structure.

**Manageability**
Sound and System engineers are not only responsible for designing and setting up systems, but they must also be able to handle unexpected situations, failures of devices, cables, power systems etc. This usually requires that systems offer tools for visualising system structures, monitoring the status of devices and connections, providing procedures for managing and fixing failures.

**Versatility**
Sound Reinforcement systems often need to be modified due to changes of location or differing requirements for certain productions in venues. For rental companies Loudspeaker Systems generally have a deeply modular character with changing configurations and scales daily.
Networked signal structures must be able to support this versatility while still maintaining reliability and the desired workflow efficiency.
The combination of these requirements for usability usually differs deeply from traditional ways of designing, assembling, and managing network systems as for example by conventional IT departments.

## 5 Technical difficulties with Ethernet - based audio networks

Ethernet is in its original design not made for transporting media signals with low bounded latency and guaranteed arrival of packets to many synchronised receivers.

Over time certain protocols were added to Ethernet for enabling a manageable Quality of Service (Qos, 802.1q), synchronisation between network nodes (IEEE 1588) and the ability to send data to multiple receivers without flooding the entire network (Multicast Protocol IGMP). These protocols have not been conceived with ProAV use cases in mind and their implementation in network devices is rarely taking ProAV requirements into respect.

### 5.1 Quality of service (QoS) / Diffserv

The introduction of Diffserv in 1998 [6] extended the header in Ethernet IP packets by a byte in which 6 bits indicate a certain priority class of traffic. This is

called 'Differentiated Service Code Point' / DSCP. According to the DSCP value, packets are prioritised by Switches in different queues over others. Prioritisation is executed with different configurable algorithms like 'Strict Priority' or 'Weighted Round Robin'.
This prioritisation is only relative and does not guarantee delivery of data in case of traffic overloads. There is also no clear agreement about which DSCP codes to use for which type of traffic so different protocols can easily get in conflict with their priorities.

DSCP always comes with significant effort for configuration of network switches because switches must be configured to treat the DSCP classes used by Audio network protocols in the right way.

## 5.2 Precision Time Protocol IEEE 1588

IEEE1588 was introduced in 2002 in its first version and has evolved since in multiple versions [7]. The Precision Time Protocol (PTP) enables network nodes to synchronise their internal time to a leader in the network. This happens by sending specific time messages between nodes and measuring the transmission delays for these messages. Nodes can then align their internal wall clock to the leader with a high accuracy. This accuracy is deeply dependent on the version of PTP used and its implementation in switches. It is also affected by other network traffic.

In **PTP V1** switches have no active role in the synchronisation itself. This has the effect that any variation in transport delays caused by traffic in switches would affect the accuracy of the PTP. The inaccuracy is highest when delays are asymmetric regarding the direction of traffic. This is the most common case in network systems distributing signals from few sources at one end to many receivers at the other end as in loudspeaker systems.

**PTP V2** was introduced in 2008 and it added clocks in switches which can adjust time stamps in packets according to the traffic delays in the switch. This vastly improves the accuracy of synchronisation.
Still this improvement depends on if the so-called Boundary or Transparent Clocks in Switches are implemented in Hardware or in Software. Hardware implementations are generally far more accurate and less affected by actual other traffic.

**IEEE 802.1as** - also called gPTP - is a part of the IEEE AVB protocol suite. It was published in 2011 and is a version of IEEE 1588 with higher precision. This is because it uses IEEE 1588 mechanisms at Layer 2 of a LAN. In 802.1as clocks are implemented in hardware at each network port and the protocol mechanisms are executed directly between network ports rather independent from higher level network traffic. This makes the synchronisation very robust and precise.

This is an overview over some relevant Audio Network protocols and their PTP version [8]:

| | |
|---|---|
| Dante: | PTP V1 |
| Q-Lan, Ravenna, AES67: | PTP V2 |
| Milan (AVB): | gPTP 802.1as |

## 5.3 Accuracy of PTP versions

In principle all versions of PTP can achieve high accuracy in the required range for professional applications. However, the accuracy practically achieved in systems very much depends on various factors, especially on the scale of the network and the actual traffic load in the network. As PTP packets are often prioritised using Diffserv, PTP can become inaccurate when Diffserv queues congest.

This has been examined in various experimental test setups [6]. In this test a network of 6 daisy-chained switches was loaded close to max. bandwidth with Video, Audio and best effort traffic. The test was performed using either PTP V1, V2 or 802.1as gPTP (AVB) synchronisation. Results are summarized in Table 2.

| Max Time Error | $1^{st}$ hop | $6^{th}$ hop |
|---|---|---|
| PTP V1, heavy traffic | N/A | ±2500µs |
| PTP V2, transparent clocks | ±0.024µs | ±4.5µs |
| PTP V2, boundary clocks | ±1µs | ±4µs |
| gPTP | ±0.035µs | ±0.063µs |

Table 2: Maximum time errors according to [9]

Test results showed significant differences in the accuracy of the different PTP versions.

**PTP V1:** The timing accuracy varies significantly with the network traffic load, and it clearly exceeds the required precision of 1us. The time reference varies after 6 hops up to 2.5ms with heavy traffic load, but even with low traffic it shows errors up to 10µs.
**PTP V2:** After 6 hops and with heavy network traffic PTP V2 shows timing errors in the range of 4 - 4.5µs. The difference between Boundary Clocks and Transparent Clocks is rather low. After 1 hop PTP V2

with Boundary Clocks shows very low errors at 24ns while using Transparent Clocks still shows 1µs of error.
**802.1as, gPTP:** Even with a high load of traffic gPTP stays well below 100ns of timing errors and the error varies only slightly between 1st and 6th network hop.

gPTP synchronized reliably across 6 hops within 3 seconds, whereas PTP with Boundary Clocks required over 3 minutes for synchronization [9].

From this test [9], gPTP 802.1as (AVB) is the only PTP candidate fulfilling the previously described technical requirements for timing accuracy described earlier in section 2.

## 5.4 Implications on sound reinforcement systems

Line source systems where individual loudspeaker elements form an array in which the dispersion is achieved by a certain relation of distance and angle of the elements and their transducers the timely relation between the elements. These inter element angles are optimized in system simulation software, such as d&b array calc or L-Acoustics Soundvision, together with electronic filtering to optimize the response of the system as seen from the audience.
Modern Line Source systems provide mechanisms and fittings for adjusting the relative position between elements within 1 mm of accuracy or less.

To illustrate the effect of such uncertainty, we use an example of a 12 element L-Acoustics K2 array, that has been optimized to cover a flat field from 6 to 60m. In this example, each loudspeaker element is associated to a single network node corresponding to an external amplified controller, delivering 4 audio channels (2 for low frequencies, 1 for mid frequencies and 1 for high frequencies). Processing includes digital crossovers, loudspeaker protection and optimization, and digital optimization of the configuration (Autofilter). Each network node has a random time offset to emulate synchronization errors. Alternative designs using 2 or 3 line source elements in parallel were simulated but leading to similar results.

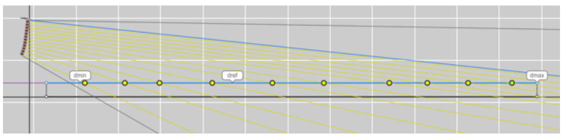


Figure 2: loudspeaker system and audience area used for synchronization evaluation tests

Figure 3 illustrates frequency response variations between unsynchronized and synchronized elements in the loudspeaker system under test, considering uniformly distributed timing offsets of ±0.5µs. That may result either from mechanical uncertainty, or network synchronization errors against a perfectly mechanically arranged and synchronized array.
The differences remain limited to ±0.5dB and are restricted to frequencies above 10kHz. This corresponds to the recommended 1µs maximum time offset. Observed deviations can be considered as inaudible.

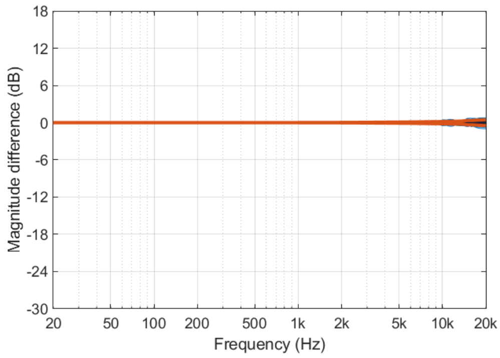


Figure 3: frequency response variations between unsynchronized and synchronized according to Figure 2 (blue: individual curves, black, median, red: 2.5 and 97.5 percentile), third octave smoothing, **±0.5µs** timing offset (uniform random distribution)

Figure 4 illustrates variation in timing offset randomly distributed between ±10µs. We see that frequency response variations start from 1kHz upwards and increase with frequency, reaching already +/- 3dB at 5kHz well above the audibility threshold of frequency response irregularities.

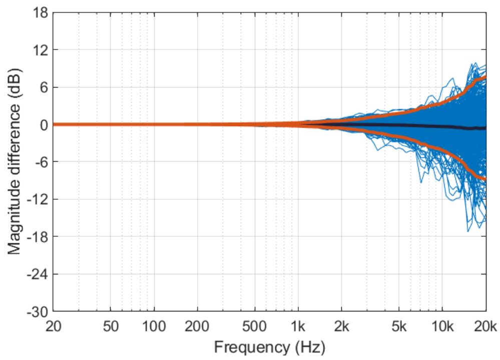


Figure 4: frequency response variations between unsynchronized and synchronized according to Figure 2 (blue: individual curves, black, median, red: 2.5 and 97.5 percentile), third octave smoothing,

**+/- 10µs** timing offset (uniform random distribution)

Relating to time offset errors observed with PTP V1 or V2, it looks that PTP V1 is not suitable for modern line sources with individually processed elements. Even in low traffic on network, time errors are reaching 10µs which may already compromise sound quality. Additional traffic may increase these timing errors among network nodes and further deteriorate sound quality.

## 5.5 Multicast Protocol IGMP

Ethernet itself natively only features data traffic from one sender to one receiver (Unicast) or from one sender to all receivers (Broadcast). Certain MAC addresses are reserved as Multicast addresses (one sender to N receivers) but such packets still need to be received by all receivers before the receiver can decide if it wants to make use of the packet.
Practically, if an audio signal shall be sent to only a few receivers in a network it must be sent via Unicast multiple times to each receiver. This is very inefficient with regards to network bandwidth, handling and resources in the transmitter. Using Ethernet Broadcast or Multicast addressing would spread the data across the entire network and could thereby cause congestion in switches and receivers.
The need for Multicast type of traffic is very high in Sound Reinforcement systems because it is a common case that one signal must be distributed to many nodes such as amplifiers or loudspeakers. The more often DSP processing in the amplifiers or powered loudspeakers is covering time alignment and equalisation of the system the higher is the demand for Multicast streams in a Sound Reinforcement system network because identical signals will be sent to many receivers.

The most common solution to this problem is the **Internet Group Multicast Protocol (IGMP)**. It enables receivers to announce their interest in certain data streams. Transmitters send packets only once, the duplication of packets to only the relevant receivers and the management of receivers in so-called Multicast Groups happens in IGMP capable routers and switches implementing IGMP snooping. IGMP was conceived with media applications in mind, esp. cable TV in buildings. It exists in 3 different versions which are backwards compatible.

## 5.6 Problems with IGMP

IGMP is a complex protocol which requires careful configuration in all switches of a network. It is often implemented in different versions and with different terminologies which is very difficult to handle for users. A frequent practical problem is that even only one wrongly configured switch can flood a network with multicast data packets and cause congestion in network segments. Many devices incorporate small network switches which don't support IGMP snooping and this can cause various issues with congestion and drop-outs.
Once such problems with IGMP occur in a network, they are hard to trouble-shoot. It often requires investigating the network traffic at the level of packet details for identifying a wrongly configured switch or node.

## 5.7 The problem of Convergence

Networked Sound Reinforcement systems require at least two different types of data traffic on the network: Audio and Control data. As described in section (3) under (Controllability) control data can have different criticality depending on the use case. It is highly desired to flow on the same port or cable as the audio so that one network cable to the device is sufficient, resp. two for redundant layouts.

The only mechanism in legacy Ethernet for managing the coexistence of highly critical audio traffic with more or less critical data traffic is Diffserv (QoS) which only provides a relative prioritisation of traffic types. There is no mechanism associated with Diffserv that would prevent an Ethernet network from being congested by too much inrushing traffic. Data traffic often comes in short bursts which can create an overload in the buffers of switches for a short period of time. This can cause two effects:

- Because the time message packets of the legacy IEEE1588 PTP protocols V1 and V2 are using the same layer of QoS as any other traffic, short overloads can affect the timing precision as shown in Figure 3.
- Bursts of control traffic can also directly congest queues in switches and cause dropouts in audio flows.

This risk of interference between audio and data traffic can be minimised by providing extensive headroom in the network bandwidth but the effect is again only statistical. To reliably avoid such interference most networked devices have separate Ethernet ports for audio network and for control. This workaround creates quite some overhead in network resources, effort in cabling and network management while it still doesn't meet the real-world requirements.

If the network for the Sound Reinforcement system extends across an entire AV system - as e.g. for a stage or a hall - it becomes very likely that other network protocols will have to be transported as well. Only for audio the most common ones would probably be Dante, Q-Lan, AES67, Ravenna and Milan. If Video is also using the same network SMPTE 2110, IPMX, NDI and/or SDVOE signals are likely to be present on the same network.
This creates a situation of traffic with 'mixed criticality' and very different bandwidth and synchronisation requirements being present on a network that offers no real management of bandwidth and traffic classes.

Again, as with Control and Audio traffic, Diffserv QoS can only provide relative priorities between these protocols and the different standards are partly competing about prioritisation. They often also use different versions and profiles of PTP which are not compatible and are only very rarely at all supported on one switch at the same time.

Configuring VLans in Switches for the different traffic types is basically useful and common best practice. Still, in most switching architectures the VLans are merely logical separations. The traffic on the different 'logical lanes' at least still shares the same buffers, queues and wires on the trunk lines between switches and this means that again QoS handles the priorities on these connections for all traffic together and PTP is most often only supported reliably in one version or profile at a time.

In practice the problem of convergence creates a very hard to oversee 'matrix of incompatibilities' in a network system that is hard to manage by users. For the IT manufacturers this is a very uncommon situation, and it is usually not at all supported in the switch products or by technical support.

For many users, system designers and manufacturers, one solution is to separate the Sound Reinforcement part of AV systems in a separate network structure due to its very high criticality for reliability and timing accuracy while at the same time requiring user friendly workflow efficiency and deep versatility. Most often this separation is relatively easy to achieve because loudspeaker systems are usually at the end of the audio signal path. They might be the largest structure in many AV systems but they are for audio rather a 'large sink' than a 'source'.

# 6 Benefits of deterministic network technologies

## 6.1 Audio Video Bridging

With respect to the requirements in professional AV systems and some other applications the IEEE 802.1 working group in 2006 started work on enhancements to Ethernet for improving its usefulness for media applications [10] [11]. This workgroup in IEEE was called Audio Video Bridging (AVB) and it aimed to achieve the following improvements to the Ethernet network layer 2:

- Very precise synchronisation between all network nodes without need for configuration and virtually independent in its performance from other traffic in the network.
- Low and bounded latency with very low jitter.
- Transport of critical Audio and Video signals in reserved streams which are not affected by other traffic in the network.
- Stream Reservation also covers requirements for Multicast because AVB Switches automatically distribute Media Streams only to registered Listeners.

These capabilities are provided by a suite of protocols which together form the AVB standard. It is the first version of deterministic Ethernet where 'deterministic' means that the arrival of a data packet is not subject to statistical relations between types of traffic but can be guaranteed in terms of time and reliability.
The core of AVB technology is that Switches play a different role in the functioning of a network. Switches are actively taking care of gPTP on their hardware ports, they reserve bandwidth for streams throughout a network structure and they shape incoming traffic on network ports for avoiding bursts and congestion. The latter is one of the most important functions of AVB in contrast to legacy Ethernet where no mechanism is controlling the amount of traffic rushing into a switch or a network.

## 6.2 Credit Based Shaping

Ethernet was originally based on the principle of a shared medium (the 'Ether') where transmitters would send data in a rather uncontrolled fashion. If data would collide and get lost the transmitter would according to certain rules send it again after a moment of time. Such a handling of traffic is for obvious reasons not appropriate for synchronous time critical traffic. Over time other protocols for continuous audio and video streams in Ethernet such as UDP and

RTP were developed but they cannot manage loss of packets.

With 802.1Qav a credit-based traffic shaper is directly introduced into Ethernet. It is a protocol that for each network port establishes queues for critical traffic and sends messages to a connected transmitter whether to send more packets or to wait. The algorithm is based on ‘credits’ built up by the flow of the incoming traffic.
This mechanism very efficiently smooths out the traffic on network ports and avoids congestions in the network. Because it is working at Layer 2 of the network all higher layer traffic can in principle benefit from it as well. Ports using 802.1Qav can for example smooth out traffic for TCP/IP, Dante or AES67 as well if the traffic shaper is supported by both sides of the connection.

Traffic shaping avoids that too much traffic is rushing into network ports which means that it prevents the root cause for data congestion and interferences in networks. This is fundamentally different from Diffserv which only aims to manage incoming traffic by priorities but cannot prevent buffers from overloading in the first place.

## 6.3 Timing and Synchronization

The technical background to IEEE 802.1as gPTP and its outstanding precision was already explained and shown in section 5. AVB utilises this precision for a certain method of synchronising media devices in a network.

Rather than letting every media device derive its media clock from the network PTP time, audio and video streams are sent with a so-called ‘Presentation Time’ in their headers. The sequence of incoming packets with each their presentation time stamp forms the trigger for the device internal Media Clock generator.

In AVB media clock is thereby decoupled from network time and can be different for every Stream and every device. This makes it possible that within one network groups of devices can form a Media Clock Domain and different Media Clock Domains can coexist without interference.

The network itself can of course not convert Clock Rates between different Media Clock Domains, this requires devices with Sample Rate Conversion.

**Network latency:** Ethernet switches work according to certain principles and agreements on the max. length of frames and the verification of packets. These principles cause a certain latency for each switch hop and this is not different for AVB. It means that in a network structure with 1Gbps 1ms of latency can be achieved across 7 network hops. 7 is a number that is commonly accepted as a realistic number of switch hops even for large systems.

## 6.4 Stream Reservation Protocol

Stream Reservation is the complementary mechanism to Traffic Shaping for making media data flow fully undisturbed through a network structure. The 802.1Qat Stream Reservation Protocol (SRP) defines mechanisms for Switches to respond to connection requests by endstations by reserving bandwidth end-to-end along the entire path through the network.

While in legacy Ethernet there are no such entities as ‘paths’ or ‘streams’, SRP introduces end-to-end connections between network clients with guaranteed bandwidth and delivery. When an endstation requests bandwidth for an audio or video Stream and this bandwidth cannot be provided it will become notified by the Switch that the connection isn’t available. This way it is not possible to build routings through the network which might be unreliable.

## 6.5 Configurable Convergence of Traffic

In an AVB network, switches can be configured to a certain portion of available bandwidth being available for Media Streams while reserving the remaining bandwidth for other traffic. The default value suggested for switches is 75% for AVB media traffic. The audio streams can not exceed these *configurable* limits so a certain bandwidth will always remain for control or other audio network flows. Unused AVB bandwidth is available for any best effort traffic This effectively enables having Control and Audio (or Video) on one port or cable without interference.
Bandwidth and Stream reservation don’t solve all existing incompatibilities between different network protocols but in the first place they provide very good and reliable convergence between control and audio for Sound Reinforcement systems. The core requirements described in section 3 and 4 can very well be fulfilled.
Practical experiences have shown that other protocols can coexist in an AVB network without problems. Dante can especially even benefit from the internal Traffic Shaping between switches because of its PTP

V1 time message packets flowing through the network with less variation in delay.

## 6.6 Multicast in AVB

The mechanism of Stream Reservation has a very beneficial side effect on the topic of Multicast. Network internally, AVB streams are always treated as Multicast but Switches will only provide them on those ports, where a Stream Reservation has been initiated. This is an inherent part of the SRP protocol and doesn't require any user configuration. It means that the problems with IGMP described in section 5 do not exist for Media Streams in AVB networks. The correct and most efficient distribution of Streams happens at any time automatically.

# 7 Limits of AVB

**Routability**

The outstanding performance of AVB comes mainly from the fact that the protocols work on network OSI layer 2 which in essence means that they are implemented more directly to hardware and are not affected by higher layer complex software processes in Switches and end nodes.

The price to pay for these advantages is that AVB itself does not work across network routers. It does not use IP addresses (only Mac addresses) and thereby it is limited to run inside a LAN structure.

This might be seen as a drawback but practically it even supports reliability because a Sound Reinforcement System with AVB is much less likely to be affected by traffic or intrusions from outside a local system.

When different systems in a venue or site shall be interconnected AVB would require devices between systems that act as Gateways. Such Gateways are most often anyway required to manage control and audio connectivity and secure against unwanted influences from the outside of a system.

**Scalability**

The protocols described here basically scale over any network bandwidth so they in principle apply to 100 Mbps as well as to 100 Gbps and they can be applied to different media such as two- wire Ethernet or fibre connections of course.

The number of hops is *in principle* not limited but every switch hop adds to the end-to-end latency which requires buffering in the end devices. This buffering is usually - as an agreement - limited to 2ms of latency and this limits the number of hops to 14 assuming a per-hop latency of 125us with some headroom.

It is important to note that this is the max. path length through switches in the network, not the max. number of switches. Practically 14 hops is not easy to reach in a loudspeaker system unless devices offer switch ports for daisy-chaining and this feature is used extensively.

**Number of Streams and Channels**

Another limit of AVB lies in the details of the Stream Reservation Protocol. SRP establishes a communication scheme between nodes and switches for reserving bandwidth and this scheme works with certain time-outs for handshakes and for devices having left the network. The mechanics of these timing parameters cause that AVB currently supports 150 streams firmly reserved in a LAN. One stream can contain up to 8ch in Milan Base Formats, extended High-Capacity Formats enable up to 60ch per stream so that still thousands of audio channels can be transported in one structure.

# 8 The role of MILAN

IEEE standards define technological principles and protocols, but they do not standardise implementations for achieving practical compatibility. AVB has in the past been implemented by several manufacturers in various flavours and with deviations that made devices often virtually incompatible.

In 2016 leading manufacturers of Sound Reinforcement systems such as d&b Audiotechnik, L-Acoustics, Meyer Sound, RCF and Adamson System Engineering started discussions about a coordinated approach for implementing AVB into AV systems to achieve full cross-manufacturer compatibility and to build a future-proof foundation for developing the standard further. This approach has led to forming a working group under the umbrella of the Avnu Alliance and the name of this workgroup as well as of the AVB implementation agreement itself is 'MILAN'[1].

Milan defines several implementation details for AVB such as:

- Audio Packet Format (AAF, 32bit)
- Channel counts in Streams (1,2,4,6,8) + optional High Capacity Formats

[1] https://avnu.org/Milan/

- Network Redundancy
- Network Management
- Media Clock Domain management
- Simple Access Control

These are mostly just agreements but the work on Milan has also triggered improvements of the related IEEE standards.

**Milan - a collaborative Ecosystem**

A fundamental requirement to the market environment for Sound Reinforcement systems is compatibility and interoperability across manufacturers, regions and specific applications. The best technology would be meaningless if it was limited to a closed proprietary island and in competition with incompatible variants.

The Milan initiative makes open standard AVB/TSN technology practically usable for high profile Sound Reinforcement applications and secures compatibility and ongoing future development in steps that are aligned with the requirements of users and markets.

## 9 Conclusions

Besides the measurable benefits in the technical performance which fulfil all the described technical requirements for Sound Reinforcement the combination of AVB protocols in the Milan application solution provides significant benefits to users and their needs.

Considering lack of legacy Ethernet and how they are intended to be fixed by PTP V1 & V2, Diffserv and IGMP, for users the biggest practical problem with legacy Ethernet based networking is the effort for configuring networks to make them work while still finally having to accept a just 'statistical' reliability with audible deviations in timing and possible failures in the signal transport.

AVB very elegantly solves the problems of Quality of Service, offers an unprecedented precision in time and synchronisation, and manages signal distribution without any need for configurations or special settings in a network. Users still must calculate bandwidth requirements for their network architecture and consider which other protocols to support but the handling of AVB itself is usually without any efforts for configuration. To users an AVB network presents itself close to plug-and-play as a very robust part of their Sound Reinforcement system. This very much changes the perception of network from a "trouble making risk" to something that "just works".

## References

[1] Hemming, Roland. "Networked Audio for Live Events." *Audio Engineering Society Conference: UK 18th Conference: Live Sound*. Audio Engineering Society, 2003.

[2] Gross, Kevin. "The Future of High-Performance Media Networking." *Audio Engineering Society Conference: 44th International Conference: Audio Networking*. Audio Engineering Society, 2011.

[3] Travis, Chris and Lesso, Paul, "Specifying the Jitter Performance of Audio Components.", 117th Convention of the Audio Engineering Society.

[4] Frandsen, Christian G., and Chris Travis. "Clean Clocks, Once and for All?." *Audio Engineering Society Convention 120*. Audio Engineering Society, 2006.

[5] Moore, Brian CJ, and Chin-Tuan Tan. "Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion." *Journal of the Audio Engineering Society* 52.9 (2004): 900-914.

[6] Blake, Steven, et al. An architecture for differentiated services. No. rfc2475. 1998.

[7] Stanton, K. B. (2018). Distributing deterministic, accurate time for tightly coordinated network and software applications: IEEE 802.1 AS, the TSN profile of PTP. IEEE Communications Standards Magazine, 2(2), 34-40.

[8] Walsh, Kieran. "Dante and AVB Networking." *Audio Engineering Society Conference: UK 24th Conference: The Ins & Outs of Audio*. Audio Engineering Society, 2011.

[9] Alon Regev & Bogdan Tenea, "Is AVB Really Much Better than Using Classic 802.1Q with SDN?" for IEEE Automotive Ethernet & IP Technology Day, 22 September 2016, Paris, France

[10] Panel, D. "An Introduction to AVB (AUDIO / VIDEO BRIDGING), the 1st TSN profile", webinar (2022).

[11] Garner, G., & Rodriguez, S. (2022). The Transport and Impact of Synchronisation in Time-Sensitive Networking, webinar.